Efficient Algorithms Beyond ℓ_2-norm.

Logistics: - Pset 2 due next Tuesday
- Pset 1 solutions will be posted soon

Recap last time.

- Efficient algo for mean estimation for $\mathcal{G}_{cov}(\sigma)$
- Writing down non-convex problem analogous to MD functional
- Only non-convex b/c mean can change
- Show (using modulus of continuity) that mean doesn't change too much, hence "that" non-convex $\Rightarrow$ approx. global opt.

$\left(\frac{1-\varepsilon}{1-3\varepsilon}\right)^2 \Rightarrow$ tight ($\varepsilon = \frac{1}{3} \Rightarrow$ arbitrarily bad local optima)

This time.

- Resilience in other norms
- Extend algo to other norms using SDP
- Grothendieck's inequality $\leftarrow$

Estimation in other norms.

- So far focused on ℓ_2: $\|\hat{\mu}-\mu\|_2$ small
- Other norms: $\|\hat{\mu}-\mu\|_1$
  - Analog of $\mathcal{G}_{cov}(\sigma)$?

Dual norm. Given a norm $\|\cdot\|$, the dual norm $\|\cdot\|_*$ is $\|u\|_* = \sup_{\|v\| \le 1} \langle u, v\rangle$.

Examples.

norm	dual	
$\|\cdot\|_2$	$\|\cdot\|_2$	$\sup_{\|v\|_2 \le 1} \langle u, v \rangle = \|u\|_2$ $\langle u, v \rangle \le \|u\|_2 \cdot \|v\|_2 \le \|u\|_2$
$\|\cdot\|_1$	$\|\cdot\|_\infty$	$\sup_{\|v\|_1 \le 1} \langle u, v \rangle = \|u\|_\infty$ $\le \|u\|_\infty \cdot \|v\|_1 \le \|u\|_\infty$
$\|\cdot\|_\infty$	$\|\cdot\|_1$	

- $\|\cdot\|_{**} = \|\cdot\|$ (finite-dimensional)
- dual norms: "make Hölder's inequality true"

$\langle u, v \rangle \le \|u\|_* \|v\|$, and this can be made tight

$$\|u\|_* = \sup_{\|v\| \le 1} \langle u, v \rangle.$$

$\|\cdot\|_p$ $\qquad$ $\|\cdot\|_q$

$$\frac{1}{p} + \frac{1}{q} = 1$$

Matrix norms:

$\|\cdot\|_F$	$\|\cdot\|_F$
$\|\cdot\|_{op}$	$\|\cdot\|_{nuc}$ ← sum of singular values ("ℓ_1")

max singular value ("ℓ_∞")

$\|v\|_{(k)}$ = sum of absolute values of k largest entries of v

what is dual norm? → $\max\left(\|u\|_\infty, \frac{\|u\|_1}{k}\right)$

$k=1$: ℓ_∞-norm: dual should be ℓ_1-norm

$k=d$: ℓ_1-norm: dual should be ℓ_∞-norm

$k=2$: sum of the largest entries

$$\sup_{\substack{\text{sum of 2 largest} \\ \text{entries of } v \le 1}} \langle u, v \rangle = \max\left(\|u\|_\infty, \frac{\|u\|_1}{2}\right)$$

$\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$

$(1,1,1) \Rightarrow 1.5$

Generalizing $\mathcal{G}_{cov}(\sigma)$.

$$\mathcal{G}_{cov}(\sigma) = \{ p \mid \|\mathrm{Cov}_p[X]\| \le \sigma^2 \}$$

$$= \{ p \mid \sup_{\|v\|_2 \le 1} v^T \mathrm{Cov}_p[X] v \le \sigma^2 \}$$

$$\mathcal{G}_{cov}(\sigma, \|\cdot\|) = \{ p \mid \sup_{\|v\|_* \le 1} v^T \mathrm{Cov}_p[X] v \le \sigma^3 \}$$

dual norm

Proposition. If $p \in \mathcal{G}_{cov}(\sigma, \|\cdot\|)$ and $r \le \frac{p}{1-\varepsilon}$, then $\|\mu(p)-\mu(r)\| \le \sqrt{\frac{2\varepsilon}{1-\varepsilon}}\cdot\sigma \approx \mathcal{O}(\sqrt{\varepsilon})\cdot\sigma.$

$\Rightarrow$ E.g. p is $(\mathcal{O}(\sigma\sqrt{\varepsilon}), \varepsilon)$-resilient in $\|\cdot\|$).

Pf. $\|\mu(p)-\mu(r)\| = \sup_{\|v\|_* \le 1} \langle \mu(r)-\mu(p), v\rangle$

used $\|u\|_{**} = \|u\|$

$$= \sup_{\|v\|_* \le 1} \mathbb{E}_r[\langle X, v\rangle] - \mathbb{E}_p[\langle X, v\rangle]$$

Bounded variance!!

$$\le \sqrt{\frac{2\varepsilon}{1-\varepsilon}}\,\sigma.$$

$\mathrm{Var}[\langle X, v\rangle]$
$= v^T \mathrm{Cov}[X]\cdot v \le \sigma^2$ ∎

Idea for general algo:

$$\min_{q_1,\dots,q_n} \sup_{\|v\|_* \le 1} \sum_{i=1}^{n} q_i \langle X_i - \mu_q, v\rangle^2$$

s.t. $\sum_i q_i = 1$, $q_i \ge 0$, $q_i \le \frac{1}{(1-\varepsilon)\cdot n}$

Need. $\sup_{\|v\|_* \le 1} v^T \Sigma v \Rightarrow$ NP-hard.

$\|u\|_1 \Rightarrow \|v\|_* = \|v\|_\infty \quad \sup_{\|v\|_\infty \le 1} v^T \Sigma v$

A set $\mathcal{C}$ is a K-approx. relaxation if:

$$\sup_{M \in \mathcal{C}} \langle M, \Sigma \rangle \le K \cdot \sup_{\|v\|_* \le 1} v^T \Sigma v \quad \forall \Sigma \succeq 0$$

$\hookrightarrow \langle M, S \rangle = \sum_{ij} M_{ij} S_{ij}$

All $M \in \mathcal{C}$ should be p.s.d.

$vv^T \in \mathcal{C} \quad \forall \|v\|_* \le 1$

$\mathcal{C} = \{ vv^T \mid \|v\|_* \le 1 \} \Rightarrow 1$-approximate relaxation

$\langle vv^T, \Sigma \rangle = v^T \Sigma v$

$(x_i - \mu_q)^T \overset{M}{\cancel{vv^T}} (x_i - \mu_q)$

$$\min_{q_1, \dots, q_n} \sup_{M \in \mathcal{C}} \sum_{i=1}^{n} q_i \langle x_i - \mu_q, v \rangle^2 \Rightarrow q_i (x_i - \mu_q)^T M (x_i - \mu_q)$$

s.t. $\sum_i q_i = 1, \quad q_i \ge 0, \quad q_i \le \frac{1}{(1-\varepsilon) \cdot n}$

Proposition. If $p^* \in \mathcal{G}_{cov}(\sigma, \|\cdot\|)$, then $\|\mu_q - \mu_{p^*}\| = O(\sqrt{K\varepsilon} \cdot \sigma)$ for ε small enough.

Proof. Exercise 4.

---

k-approx. relaxation for ℓ_1/ℓ_∞-norm

$$\max \quad v^T \Sigma v$$
$$\text{s.t.} \quad \|v\|_\infty \le 1$$

$$M = vv^T$$

$$\max \quad \langle vv^T, \Sigma \rangle$$
$$\text{s.t.} \quad \cancel{\|v\|_\infty \le 1}$$

$$M_{ii} \le 1 \quad \forall i$$

$$v_i^2$$

$$\text{rank}(M) = 1$$
$$M \succeq 0$$

$v =$

$vv^T =$

$\| \ \|_{n \times n} = n$

$$\max \quad \langle M, \Sigma \rangle \qquad (\dagger)$$
$$\text{s.t.} \quad \text{diag}(M) = 1, \quad M \succeq 0$$

$\cancel{\text{rank}(M) = 1}$,

semidefinite program

Thm (Grothendieck) If $\Sigma \succeq 0$, then

$$(\dagger) \le \frac{\pi}{2} \cdot \sup_{\|v\|_\infty \le 1} v^T \Sigma v .$$

$$(\Rightarrow \; k = \frac{\pi}{2} \text{ - relaxation})$$

entry-wise

Pf. (1) $\max_{\|v\|_\infty \le 1} v^T \Sigma v \ge \frac{2}{\pi} \cdot \max_{\substack{X \succeq 0 \\ \mathrm{diag}(X)=1}} \langle \Sigma, \arcsin[X] \rangle$

$$\ge \frac{2}{\pi} \cdot \max_{\substack{X \succeq 0 \\ \mathrm{diag}(X)=1}} \langle \Sigma, X \rangle$$

(2) $\arcsin[X] \succeq X$

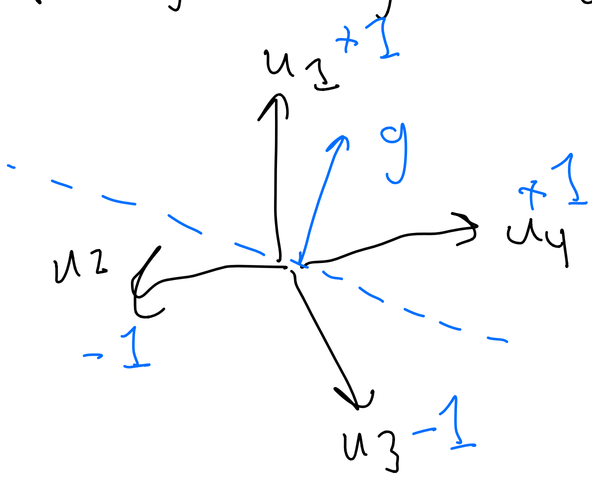
(1) + (2)

Given X s.t. $X_{ii} = 1$, $X \succeq 0$

Goal: Construct $p(v)$ s.t. $\mathbb{E}_{v \sim p}[v^T \Sigma v] = \frac{2}{\pi} \langle \Sigma, \arcsin[X] \rangle$

$$X_{ij} = \langle u_i, u_j \rangle, \quad \|u_i\|_2 = 1 \quad b/c \quad X_{ii} = 1$$

$$v_i := \text{sign}(\langle u_i, g \rangle), \quad g \sim N(0, I)$$



$X \Rightarrow u \Rightarrow$ (random) v

$$\mathbb{E}_{v \sim p}[v^T \Sigma v] = \langle \Sigma, \mathbb{E}_{v \sim p}[v v^T] \rangle$$

$$\mathbb{E}[v_i v_j]$$
$$= \mathbb{E}[\text{sign}(\langle u_i, g \rangle)\, \text{sign}(\langle u_j, g \rangle)]$$
$$= 2\,\mathbb{P}[\text{sign}(\langle u_i, g \rangle) = \text{sign}(\langle u_j, g \rangle)] - 1$$

"How large is this angle"

$\langle u_i, u_j \rangle \rightarrow \arcsin(\langle u_i, u_j \rangle)$
$= \arcsin(X_{ij})$

$\Rightarrow \mathbb{E}[v_i v_j] = \arcsin(X_{ij})$.

$\arcsin[X] \succeq X$

$$\arcsin(z) = \sum_{n=0}^{\infty} \frac{(2n)!}{(2^n \cdot n!)^2} \frac{z^{2n+1}}{2n+1}$$

$$= z + \frac{z^3}{6} + \cdots$$

$$\arcsin[X] = X + \frac{1}{6} X^{\odot 3} + \cdots$$

$X^{\odot k} \succeq 0$
$\forall k \geq 1$

<u>Claim.</u> If $X \succeq 0$, then $\underline{X^{\odot 3} \succeq 0}$

<u>Lemma.</u> If $X, Y \succeq 0$
then $X \odot Y \succeq 0$

Gaussian Cov. M

Cov. $X^{\odot 2}$

x_i, y_i

<u>Pf.</u> $X \otimes Y \in \mathbb{R}^{n^2 \times n^2}$

$$(X \otimes Y)_{ii', jj'} = X_{ij} Y_{i'j'}$$

$\Rightarrow$ eigenvalues of $X \otimes Y$ are products of eigenvalues of X, Y individually

$\Rightarrow X \otimes Y \succeq 0$ if $X, Y \succeq 0$

$X \odot Y$ principal submatrix of $X \otimes Y$. $\square$

__Example.__ Distribution learning

Let π^* be dist^n on $\{1, \dots, m\}$

Let $\tilde{\pi}$ be ε-corrupted version of π

__Goal.__ Recover $\hat{\pi}$ s.t. $TV(\hat{\pi}, \pi^*)$ small.

Dist^n of k-tuples over $\{1, \dots, m\}$

- w.p. $1-\varepsilon$: all k i.i.d. from π^*
- w.p. ε: chosen by adversary

$TV((p^*)^k, \tilde{p}) \leq \varepsilon$

Goal: Recover p^* up to small TV error.

Think of p^* as element of $[0,1]^m$

Think of k-tuples as elements of $[0,1]^m$

$m = 4$

$k = 3$

$(1,3,3) \Rightarrow \left[\frac{1}{3}, 0, \frac{2}{3}, 0\right]$

$(p^*)^k$: dist^n over $[0,1]^m$

mean of dist^n: p^*

TV distance $\Rightarrow \frac{1}{2} \cdot \ell_1$ distance

Generally: $\sup_{\|v\|_\infty \leq 1} v^T \Sigma v \leq \frac{1}{k}$

$\Rightarrow \quad \Theta\left(\sqrt{\frac{\varepsilon}{K}}\right)$ vs. $\Theta(\varepsilon)$

(interesting if $\varepsilon > \frac{1}{K}$)